Seq2seq (Sequence to Sequence)

Seq2seq е архитектура на невронски мрежи која користи два главни компоненти, енкодер и декодер, за трансформирање на една секвенца во друга. Енкодерот ја прима влезната секвенца и ја претвора во фиксна должина контекстен вектор. Декодерот потоа го користи овој контекстен вектор за да генерира излезна секвенца. Оваа техника е особено корисна за задачи каде влезните и излезните секвенци можат да имаат различни должини, како што се превод на јазик, резимирање текст или генерирање на опис на слики.

Seq2seq + Attention

Seq2seq со внимание е надградба на основната Seq2seq архитектура која вклучува механизам на внимание. Овој механизам им овозможува на декодерот да пристапи до сите контекстни вектори генерирани од енкодерот, наместо да се потпира само на еден контекстен вектор. Механизмот на внимание динамично доделува тежини на различни делови од влезната секвенца, овозможувајќи декодерот да се фокусира на најрелевантните делови за тековниот чекор на декодирање. Ова резултира со подобрени перформанси, особено за задачи со долги и комплексни влезни секвенци.

Bagging (Bootstrap Aggregating)

Bagging (Bootstrap Aggregating) е техника во машинско учење која се користи за средување на варијансата и намалување на пренатрупувањето (overfitting) на моделите. Главната цел на Bagging е да се подобрат перформансите на предвидувањето со создавање на повеќе модели, секој од кои е трениран на различно подмножество од тренинг податоците. Сите овие модели потоа се комбинираат за да се добие финалната предвидувачка способност.

Boosting

Boosting е уште една техника на ансамбл учење која има за цел да го подобри перформансата на предвидувањата. Во противност на Bagging, каде секој модел се тренира независно, Boosting методите ги тренираат моделите последователно. Секој следен модел се тренира за да ги корегира грешките на претходниот модел. Ова овозможува постепено подобрување на учењето и подобрување на крајните предвидувања.

K-means

K-means е метод во машинско учење што го користи за групирање на податоците во K кластери врз основа на сличноста помеѓу нив. Целта на K-means е да се најдат K центроиди кои го минимизираат варијансата во секој кластер, создавајќи компактни групи од податоци.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN е алгоритам за кластерирање во машинско учење што се фокусира на густина. Со DBSCAN, целта е да се идентификуваат области во податочниот простор каде што има густа концентрација на точки (примероци), што претставува кластери. Овој алгоритам е особено ефикасен за откривање на кластери со различни облици и големини, а исто така идентификува и изолирани точки кои се сметаат за шум.

Во DBSCAN, како и во K-means, нема целна колона (target column) која се предвидува, туку целта е само да се идентификуваат групите (кластерите) и шумот во податоците.

Decision Tree

Decision Tree е алгоритам во машинско учење кој создава структурирано дрво на одлуки, каде секој чвор претставува тестирање на одреден атрибут од податоците. Се користи за класификација и регресија на податоци и е особено корисен кога се бара да се истражат и одредат јасни правила за одлучување во податочните сетови. На пример, во банкарството може да се користи за да се класификуваат клиентите според нивната веројатност за враќање на кредит.

Random Forest

Random Forest е тип на ансамбл метод во машинско учење кој создава голем број од дрвета на одлуки и ги комбинира за да се постигне подобра предвидливост и стабилност. Секое дрво во случајната шума се тренира на различно обучувачко множество и избира случајни подмножества од атрибутите, што го прави отпорен на пренатренирање. Random Forest е погоден за комплексни задачи на класификација и регресија, како што се предвидување на потрошувачките навики во онлајн трговија или класификација на медицински дијагностички податоци.

Accuracy (Точност)

Точноста мери колку моделот успешно предвидува вистинските вредности од целиот набор на податоци. Оваа метрика е важна кога сите класи се третираат како еднакво значајни и кога е важно да се добие општа представа за успешноста на моделот.

Recall (Позив)

Позивот (Recall) се фокусира на способноста на моделот да открие сите вистински позитивни случаи. Во суштина, оваа метрика мери колку добро моделот ги идентификува сите реални позитивни инстанци, без да ги занемарува.

Precision (Прецизност)

Прецизноста (Precision) мери колку од предвидените позитивни инстанци на моделот се вистински позитивни. Оваа метрика е важна за да се обезбеди дека моделот не прави многу грешки кога ги класифицира инстанците како позитивни.